



Extraction et analyse de l'impact émotionnel des images

Syntyche Gbehounou, François Lecellier, Christine Fernandez-Maloigne

► To cite this version:

Syntyche Gbehounou, François Lecellier, Christine Fernandez-Maloigne. Extraction et analyse de l'impact émotionnel des images. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00656493

HAL Id: hal-00656493

<https://hal.science/hal-00656493>

Submitted on 17 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Extraction et analyse de l'impact émotionnel des images

Syntyche Gbèhounou

François Lecellier

Christine Fernandez-Maloigne

Département SIC du Laboratoire XLIM , UMR CNRS 6172
Université de Poitiers

Bât. SP2MI, Téléport 2, Bvd Marie et Pierre Curie, BP 30179
86962 Futuroscope Chasseneuil Cedex France

{syntyche.gbèhounou, francois.lecellier, christine.fernandez}@univ-poitiers.fr

Résumé

Cet article propose une méthode d'extraction de l'impact émotionnel des images à partir de descripteurs récents.

Très souvent, on associe les émotions à l'expression du visage, mais nous avons décidé de ne pas faire de cette information la principale information émotionnelle des images naturelles, qui en général ne contiennent pas de visages. Nous avons ainsi effectué nos tests sur une base diversifiée, construite à partir d'images à faible contenu sémantique.

La complexité des émotions a été prise en compte en intégrant, au processus de classification, les résultats de tests psycho-visuels que nous avons mis en place. Vingt cinq observateurs ont participé aux tests. Ils ont évalué la nature et la puissance des émotions ressenties.

Nous avons choisi un réseau de neurones multicouches pour la classification. Le taux de réussite moyen obtenu lors de la classification est de 56,15% ; ce qui est encourageant au regard des résultats de la littérature.

Mots Clef

Émotions, réseaux de neurones, tests psycho-visuels, couleurs

Abstract

This paper proposes a method to extract the emotional impact of images.

Emotions are often associated with facial expressions, but we decided not to consider this feature as first emotional characteristic of natural images, which, in general, does not contain faces. Using this statement, our tests have been done on a new image database composed of low semantic diversified images.

The complexity of emotion modeling was considered in classification process through psycho-visual tests. The twenty five observers assessed the nature and the power of emotions felt.

We used an artificial neural network for classification. The average success rate is 56,15% ; that is really relevant regarding the equivalent results in the literature.

Keywords

Emotions, Artificial Neural Network, psycho-visual tests, colors.

1 Introduction

Les images sont de réels vecteurs émotionnels. C'est pourquoi elles sont le support privilégié des médias, notamment dans un cadre publicitaire. Nous comprenons aisément que c'est l'un des moyens les plus performants pour véhiculer un message. Essayer d'identifier l'émotion d'une image est une tâche ambitieuse. Différentes informations contenues dans une image (couleurs, contenu, textures, ...) peuvent être des vecteurs émotionnels.

Mais beaucoup d'*a priori* existent sur les liens entre les images et leur contenu émotionnel notamment sur les couleurs et leur impact. De plus l'émotion d'une image est fortement liée à son contenu puisque certains lieux (plage, cimetière, ...) nous renvoient à des émotions différentes. D'autres facteurs encore plus complexes que le contenu ou la couleur globale entrent en jeu dans l'interprétation émotionnelle d'une image. Il s'agit notamment de tous les facteurs culturels et personnels. Ces facteurs sont d'autant plus difficiles à prendre en compte puisqu'ils ne pourraient être décrits par des descripteurs objectifs.

L'émotion est une réaction complexe qui engage à la fois le corps et l'esprit. Elle est déclenchée par une évaluation de la signification personnelle du sens de ce qui se produit dans la situation.

L'émotion peut être définie comme étant une réaction affective transitoire d'assez grande intensité, habituellement provoquée par une stimulation venue de l'environnement. On retrouve beaucoup d'autres définitions de l'émotion en fonction des différentes écoles. En effet, le concept d'émotion est utilisé de diverses manières selon qu'il est envisagé en référence à l'aspect stimulus, à l'expérience subjective, à une phase d'un processus, à une variable intermédiaire ou à une réponse. Cependant, globalement, on distingue dans la littérature deux théories psychologiques principales de

l'émotion :

- Les théories des émotions de base ;
- Les théories de l'évaluation.

Dans les théories des émotions de base dont les pères sont entre autres Darwin, Ekman, Izard, Plutchik [16, 17], des émotions dites basiques ou fondamentales sont énumérées. Il faut noter que seulement 5 émotions de base leur sont communes (la tristesse, la colère, la joie, le dégoût et la peur).

Dans les théories de l'évaluation [16] aussi appelées "théories de l'appraisal", les émotions sont définies comme un ensemble limité dans le temps de changements d'états en réponse à l'évaluation d'un stimulus externe ou interne particulièrement pertinent pour l'organisme .

Quelle que soit la théorie psychologique, on retrouve généralement deux classifications des émotions [2] :

- L'approche catégorielle ou discrète ;
- L'approche dimensionnelle.

Dans l'approche catégorielle, les processus émotionnels peuvent être expliqués par un ensemble d'émotions basiques ou fondamentales, qui seraient innées et communes à toute l'espèce humaine. Il n'y a pas de consensus quant au nombre et à la nature de ces émotions dites fondamentales [2].

Les modèles de l'approche dimensionnelle se différencient de ceux de l'approche catégorielle par le fait que les émotions résultent d'un nombre fixé de concepts représentés dans un espace multidimensionnel [2] .

Les dimensions peuvent être un axe de plaisir, d'éveil, de puissance. Ces dimensions varient en fonction des besoins du modèle. Le modèle le plus utilisé est celui de Russell avec les dimensions valence et activation (Valence-Arousal) :

- **La valence** représente la manière dont se sent une personne quand elle regarde, par exemple, une image. Cette dimension varie du positif au négatif et permet de distinguer les émotions négatives et agréables.
- **L'activation** correspond au niveau d'excitation corporel.

L'avantage de ces modèles est de permettre de représenter un très grand nombre d'émotions. Elles sont représentées dans un espace bidimensionnel dont les dimensions varient d'une information trop présente à pas assez. Certaines émotions peuvent être confondues (la peur et la colère par exemple) ou non représentées du tout (entre autre la surprise) dans un modèle bidimensionnel de type valence/éveil.

Dans la littérature, beaucoup de travaux se basent sur une modélisation discrète des émotions par exemple ceux de Wei et al.[20], Paleari et Huet[15], Kaya et Epps[6].

Plusieurs approches ont été développées pour l'extraction des émotions des images. L'une des techniques retrouvée le plus souvent est la détection de visages. Beaucoup de travaux reposent sur l'expression des visages pour déduire l'émotion contenue dans les images. Une émotion est alors associée à des traits du visage (sourcils, lèvres entre autres).

L'autre grande famille de travaux est celle de la détection des émotions basée sur les caractéristiques de l'image [20, 14, 10, 2]. On retrouve le plus souvent les caractéristiques suivantes :

- Couleurs ;
 - Textures ;
 - Formes présentes (détection de visage, d'animaux ...).
- Lucassen et al. [10] par exemple, définissent quatre émotions relatives aux textures qui sont :
- Chaud-Froid ;
 - Féminin-Masculin ;
 - Dur-Doux ;
 - Lourd-Léger.

À partir des résultats obtenus de l'analyse de la variance et des études précédentes faites par Ou et al.[14] (dans lesquelles on retrouve des similarités dans les échelles émotionnelles), Lucassen et al. [10] ont mis en place des fonctions permettant d'avoir une explication des variances sur l'échelle de l'émotion couleur. Ils ont établi entre autres des liens entre leurs différentes émotions et les paramètres couleur L^* , c^* et h^* .

Toujours dans le principe d'extraction de caractéristiques, certains auteurs, par exemple Wang et Yu [19], se sont intéressés à la description sémantique des couleurs. À partir de cette information, ils associent à une image une sémantique émotionnelle.

Enfin, l'extraction des formes permet entre autres de pouvoir, là encore, exploiter les expressions du visage (d'êtres humains ou d'animaux) éventuellement contenues dans l'image. Dans certains travaux, les différentes lignes contenues dans l'image et leur orientation sont prises en compte. Par exemple Dellandréa et al. [2], évoquent l'idée selon laquelle des lignes obliques communiqueraient du dynamisme et de l'action alors que des lignes horizontales ou verticales du calme et de la relaxation.

Au vu de la littérature, nous avons choisi de nous inscrire dans cette famille d'approches d'extraction de caractéristiques. Pour intégrer les différents aspects difficiles à modéliser (influence culturelle ou encore vécu personnel), nous avons décidé de mettre en place des tests psychovisuels afin de constituer une référence à notre classifieur d'images par les émotions qu'elles sous-tendent.

Après cette rapide introduction, l'article est organisé de la façon suivante. Le cadre de travail est exposé dans la partie 2 ; nous y décrirons notre base d'images et les tests psychovisuels mis en œuvre. La partie 3 est consacrée aux différents descripteurs utilisés pour la reconnaissance des émotions. Le processus de classification est détaillé dans la partie 4. Enfin, nous ferons une synthèse de notre étude et nous donnerons des perspectives à nos travaux dans la partie 5.

2 Cadre de travail

2.1 Base d'images

Puisqu'aucune base de tests d'images universelle n'existe encore dans le domaine de l'extraction de l'impact émotionnel, il est très important de pouvoir constituer la base

la plus diversifiée possible.

Notre base d'images a ainsi été constituée à partir d'images libres de droit et gratuites dans sa grande majorité. Elle est diversifiée et contient 218 images comprenant des paysages, des animaux, des personnages, des aliments, des bâtiments, comme l'illustre la figure 1. Actuellement, on re-



Figure 1: Quelques images de la base

trouve dans la base plusieurs images identiques par rapport au contenu mais différentes puisqu'ayant subi une transformation particulière (transformation géométrique, modification de la balance des couleurs). Même si ces opérations modifient dans certains cas le caractère "naturel" de l'image, elles nous permettent d'évaluer le comportement de certains descripteurs classiques de traitement d'images que nous avons choisis.

Pour associer à chaque image une "vérité terrain" quant à son contenu émotionnel, nous avons choisi de mettre en place des évaluations psycho sensorielles significatives.

2.2 Tests psycho-visuels et leur exploitation

Vingt cinq observateurs ont participé volontairement, en fonction de leurs disponibilités et gratuitement aux différents tests ; 28% étaient des femmes et 72% des hommes. La moitié était âgée de 18 à 24 ans.

La stratégie des tests est intimement liée au modèle d'émotions choisi. Nous avons décidé de travailler sur un modèle dimensionnel dont les dimensions sont :

- La nature de l'émotion ;
- La puissance de l'émotion.

Nous avons mis en place deux séries de tests effectuées à une semaine d'intervalle. Les observateurs ont évalué la nature et la puissance de l'émotion suscitée par l'image. Notre modélisation émotionnelle est équivalente au modèle Valence/Éveil (en encore Valence/Arousal) dans lequel, la valence permet de distinguer les émotions négatives et agréables tandis que l'activation, variant de faible à fort, permet de définir le niveau d'excitation corporelle. Comme on peut le voir sur la figure 2, pour définir la nature de l'émotion, les utilisateurs avaient le choix entre "Négative", "Neutre" ou "Positive". Quant à la puissance, elle variait linéairement de faible à forte.

Dans chacune des deux séries de tests, il y avait 24 images. Les observateurs disposaient de 8 secondes pour noter chaque image. Si le temps paraît aussi court c'est tout simplement pour augmenter les chances de recueillir des émotions primaires et non des émotions qui découlent d'une



Figure 2: Application de tests d'évaluation

éventuelle interprétation trop poussée de la sémantique des images.

Dans les deux séries, certaines images étaient identiques mais présentées avec un traitement différent d'une série à l'autre, voire au sein de la même série, ou encore même présentées avec deux traitements différents dans une même série. Les deux traitements appliqués aux images sont :

- Le changement de la dynamique des couleurs ;
- La rotation.

Même si ces traitements modifient le caractère naturel de certaines images, ils nous permettent de pouvoir tester l'efficacité de certains descripteurs.

Dans les images des séries de tests, il y avait très peu de visages humains. C'était une façon pour nous d'éviter l'interprétation des émotions liées à ces derniers.

L'émotion est d'autant plus complexe à modéliser qu'un simple détail sur une image peut tout changer. Ceci est illustré par les images de la figure 3. La présence de la barrière sur l'image 3(b) change le taux de classification de cette image lors des tests. Pour un nombre significatif de nos observateurs la barrière est synonyme d'interdiction de franchissement ; ce qui a donc changé l'émotion qui passe de 96% à seulement 65% d'impression positive.

La rotation appliqué sur l'image 3(d) modifie aussi le taux de classification par les observateurs.

Les résultats des tests ont servi en partie pour constituer une référence au classifieur. Une partie a servi pour l'apprentissage. Seules les images ayant un taux de classification dans une catégorie d'émotion (neutre, négative ou positive) supérieur d'au moins 10% aux autres dans cette catégorie, sont considérées comme classées dans cette dernière. S'il arrivait qu'une image soit classée à 50% dans 2 classes (cette situation ne s'est présentée dans aucun des tests effectués), elle ne serait pas conservée pour l'exploitation dans le processus de classification.

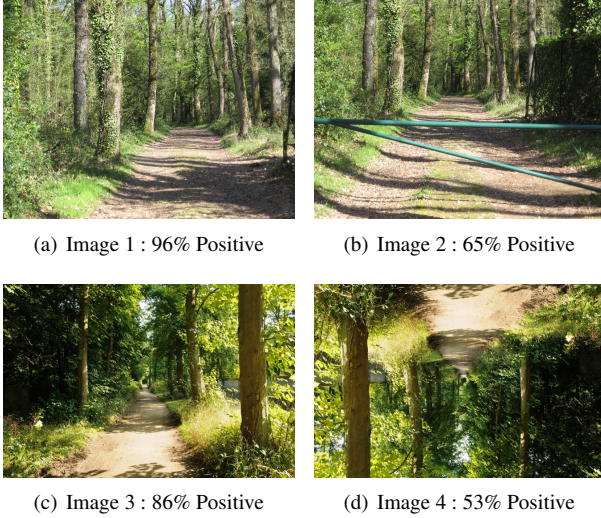


Figure 3: Puissance de l'émotion sur des images naturelles (a)-(c) et transformées (b)-(d)

3 Descripteurs utilisés pour la reconnaissance des émotions

Passons maintenant à l'extraction des paramètres qui nous ont permis d'envisager une classification authentique des images selon les "vérités terrain" données par les tests psycho-visuels.

Les couleurs sont les premières caractéristiques discriminantes des émotions. Les différentes hypothèses qui ressortent des études bibliographiques sur le lien entre les émotions et les couleurs concernent le plus souvent le rouge, le vert, le bleu et le jaune. Par exemple, selon Daniel Beresniak dans [1], le rouge est une couleur vivante et excitante, par contre la combinaison rouge+gris provoquerait un sentiment tragique. Le jaune est la couleur la plus gaie, la plus claire, rayonnante et jeune. C'est aussi une couleur tonique et éclatante. La couleur la plus dynamique selon lui c'est l'orange. Cette couleur combine la gaité du jaune et l'action du rouge. Le bleu est une couleur profonde et mystique qui appellerait au calme.

En effet, les couleurs correspondent principalement à une certaine interprétation d'une sémantique liée à des situations, à des phénomènes mais aussi à la culture. Elles ne sauraient donc être les seules caractéristiques discriminantes des émotions.

Les textures ont aussi une place importante dans le ressenti émotionnel. Par exemple une grille, quelle que soit sa couleur, a une sémantique d'enfermement ; ou encore, le métal ne fait pas le même effet que ce soit au toucher ou à la vue qu'un brin d'herbe. Enfin les descripteurs liés à la forme et aux contours des objets de l'image sont également significatifs. Nous avons aussi supposé que par leur précision, ces descripteurs encoderaient des informations de haut niveau intéressantes pour l'extraction des émotions.

3.1 Couleurs

Pour identifier les différentes couleurs, nous avons utilisé une segmentation couleur par croissance de régions [4]. Pour initialiser les germes de la croissance de régions, nous avons procédé à une analyse d'histogramme en niveaux de gris.

L'analyse d'histogramme a été faite sur l'image en niveaux de gris afin de gagner en temps de calcul dans la recherche des zones homogènes. Les germes considérés sont les maxima de l'histogramme en niveaux de gris. Si cette détection des germes a été faite en niveaux de gris, la croissance quant à elle s'est faite dans l'espace couleur CIE Lab pour minimiser les problèmes d'apparition de fausses couleurs au moment de la comparaison, par la distance euclidienne, entre la couleur du pixel à agglomérer et la couleur moyenne de la région croissante.

Nous n'avons conservé que la couleur moyenne des différents régions.

3.2 Textures

La caractérisation des textures a été faite sur les images en niveaux de gris. Nous avons utilisé deux types de descripteurs pour caractériser les textures.

Le premier descripteur est l'énergie des caractéristiques de Gabor basée sur la décomposition en bancs de filtres de Gabor [5]. Dans le domaine spatial, un filtre de Gabor est le produit d'une sinusoïde complexe et d'une enveloppe gaussienne.

Un filtre de Gabor 2D est défini en continu par la fonction $g_{\lambda, \Theta, \varphi}(x, y)$ de la façon suivante :

$$g_{\lambda, \Theta, \varphi}(x, y) = \cos \left(2\pi \frac{x'}{\lambda} + \varphi \right) \exp \left(-\frac{(x'^2 + \gamma^2 y'^2)}{2\sigma^2} \right), \quad (1)$$

avec $x' = x \cos \Theta + y \sin \Theta$ et $y' = y \cos \Theta - x \sin \Theta$.

Pour calculer les caractéristiques de Gabor, nous avons considéré 12 angles différents $\Theta \in [0, \pi[$ tous les $\frac{\pi}{12}$ et deux phases différentes $\varphi \in \{0, -\frac{\pi}{2}\}$ (0 pour le cas symétrique et $-\frac{\pi}{2}$ pour le cas anti symétrique). Ce qui revient à un total de 24 filtres.

Nous avons considéré une gaussienne isotrope ($\gamma = 1$), d'écart type $\sigma = 0,56\lambda$. Ce dernier choix est motivé par les propriétés des cellules du cortex visuel qui peuvent être modélisées par le filtre de Gabor comme le soulignent Grigorescu et al. [5].

L'énergie des caractéristiques de Gabor est la combinaison des résultats des 12 filtrages pour chacune des phases considérées.

Les coefficients Wave Atoms sont le second type de descripteurs texture que nous avons utilisés. Ils sont basés sur la décomposition en Wave Atoms introduite par Demanet et Ying [3].

Les Wave Atoms sont en première approximation une variante de paquets d'ondelettes 2D avec une longueur d'onde d'échelle parabolique. C'est la transformée la plus parcimonieuse pour la description des textures localement

oscillantes et la seule actuellement pour laquelle un tel résultat est disponible [3].

Comme dans toutes les décompositions multi-échelles (les ondelettes par exemple), on y retrouve plusieurs informations de différents niveaux.

Avec la transformée en Wave Atoms, à chaque échelle, le paramètre d'orientation est implicite et varie de manière plus fine qu'avec d'autres transformées (ondelettes par exemple). Le nombre de coefficients par orientation dépend du niveau de décomposition.

Avant d'appliquer la décomposition en Wave Atoms, nous avons redimensionné toutes les images en 256*256 en prenant soin de faire du « zero padding » si nécessaire. Avec cette taille d'image, nous obtenons 5 niveaux de décomposition. Nous n'avons conservé que l'échelle 4 qui nous offre un compromis entre le niveau de description et la parcimonie. Elle est composée de 91 orientations ; chaque orientation ayant $2^4 * 2^4$ soit 256 coefficients.

3.3 Descripteurs locaux et globaux

Nous avons utilisé les descripteurs locaux SIFT introduits par Lowe [8, 9] qui ont montré leur efficacité dans de nombreux travaux.

L'image est résumée par un ensemble de vecteurs de caractéristiques locales invariants :

- à la translation ;
- au changement d'échelle ;
- à la rotation ;
- aux changements d'éclairage ;
- aux projections affines ou 3D.

Nous avons également utilisé les descripteurs globaux « GIST » qui nous ont permis d'avoir une représentation très faible dimension. Ces descripteurs ont été introduits par Oliva et Torralba en 2001 [13]. Ils permettent de représenter la structure dominante spatiale de la scène à partir d'un ensemble de dimensions perceptives (la naturalité de la scène, son ouverture, sa rugosité, son expansion, sa robustesse). Dans un souci de réduction de dimension, ces 5 dimensions « perceptives » ne sont pas calculées mais plutôt estimées à partir des informations spectrales et des informations grossièrement localisées.

Les descripteurs « GIST » sont les seules caractéristiques qui ont été directement calculés sur les images couleur. Tous les autres descripteurs ont été calculés sur les images converties en niveaux de gris.

4 Classification des images

La classification des images est la dernière étape de notre processus d'extraction de l'impact émotionnel des images. Cette étape comporte deux phases :

- **Une phase de réduction des dimensions** : cette phase est primordiale car après les calculs des différentes caractéristiques, nous obtenons des vecteurs de dimensions importantes qui ne peuvent être directement exploités.
- **Une phase de discrimination des attributs** : la discrimination se fait à l'aide d'un classifieur, dans notre cas il

s'agit d'un réseau de neurones multicouches. Le nombre de couches dépendra des dimensions finales des informations à discriminer, suivant la catégorie des caractéristiques.

4.1 Réduction de dimensions des vecteurs de caractéristiques

Après l'extraction des caractéristiques, comme décrit dans le paragraphe précédent, nous disposons d'un nombre important d'informations, comme on peut le voir dans le tableau 1, qui donne la taille moyenne des vecteurs d'attributs par catégorie de descripteurs et par image. L'étape de

Catégories de descripteurs	Nombre moyen de descripteurs	Taille des vecteurs
Couleurs	151	3
Gabor	47 648	12
GIST	960	1
SIFT	469	128
Wave Atoms (Échelle 4)	91	256

Tableau 1: Taille moyenne par image des vecteurs de caractéristiques avant la réduction de dimension

réduction des dimensions est donc très importante si on veut pouvoir utiliser un classifieur de type réseau de neurones sans augmenter considérablement les temps de calcul.

Les différents vecteurs d'attributs ont été normalisés (normalisation L2) avant la phase de réduction de dimensions. Plusieurs solutions s'offrent à nous pour réduire les dimensions mais elles ne sont pas toutes exploitables dans le cas de l'extraction des émotions. Dans notre cas, par exemple une Analyse par Composantes Principales (ACP) n'est pas très facilement exploitable. Certains descripteurs notamment les SIFT et les couleurs n'offrent pas le même nombre de caractéristiques sur toutes les images. Aussi, faire une ACP nous obligerait à avoir un espace de référence et donc à choisir des images de référence (ce choix étant délicat).

La solution que nous avons donc adoptée pour réduire les dimensions avant l'étape de classification est une quantification vectorielle avec un algorithme de type *k-means*.

La relation entre le nombre de centroïdes et le nombre de descripteurs est donnée par :

$$k = \sqrt[4]{N * d}, \quad (2)$$

avec k le nombre de centroïdes, N le nombre de descripteurs et d la taille du vecteur de caractéristiques. L'ensemble des centroïdes constitue donc k vecteurs de dimension d.

Les dimensions des centroïdes sont récapitulées dans le tableau 2 pour chaque catégorie de descripteurs.

Une fois ces centroïdes calculés, pour chaque image et suivant chaque catégorie de descripteurs, nous avons effectué une quantification vectorielle. Pour la discrimination lors de

Catégories de descripteurs	Nombre de centroïdes	Taille des vecteurs
Couleurs	13	3
Gabor	72	12
GIST	15	1
SIFT	41	128
Wave Atoms (Échelle 4)	32	256

Tableau 2: Configuration des différents centroïdes obtenus après l'algorithme des *k-means*

la classification nous avons utilisé les histogrammes répartition.

Comme les vecteurs de caractéristiques, les histogrammes sont normalisés en utilisant une norme L2 avant la classification. Une fois les histogrammes calculés on peut passer à l'étape de classification proprement dite.

4.2 Discrimination des images

Comme nous l'avons évoqué dans l'introduction, les tests ont été faits pour constituer une référence à nos classifications finales. Une partie des tests (59%) a servi pour l'apprentissage du système de classification et l'autre partie comme séquence de généralisation. Une première étape d'analyse des résultats a été réalisée pour déterminer les différents pourcentages de classification des images dans les trois classes considérées (émotions négatives, neutres et positives).

La phase d'apprentissage s'est déroulée sans erreur sur tous les 5 types de descripteurs étudiés. Nous avons effectué une classification marginale suivant chaque catégorie de descripteurs.

Le nombre moyen d'itérations sur les 5 types de descripteurs est de 10 000. Ce nombre élevé est lié à la grande diversité des images d'une même catégorie d'émotions.

On peut observer cette grande diversité sur les histogrammes de répartition des coefficients Wave Atoms (figure 5) des images de la figure 4.

4.3 Quelques résultats

Les taux de classification que nous présentons (tableau 3) sont ceux obtenus suivant chaque type de descripteurs, sur 19 images qui correspondent à celles des tests psychovisuels qui n'ont pas servi pour l'apprentissage. Le taux

Catégories de descripteurs	Nombre d'erreurs	Taux de classification
Couleurs	8	57,89%
Gabor	7	63,15%
GIST	11	42,1%
SIFT	8	57,89%
Wave Atoms (Échelle 4)	6	68,42%

Tableau 3: Taux de réussite de classification obtenus



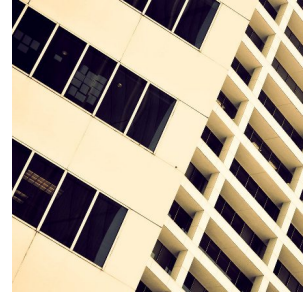
(a) Image positive 1



(b) Image positive 2

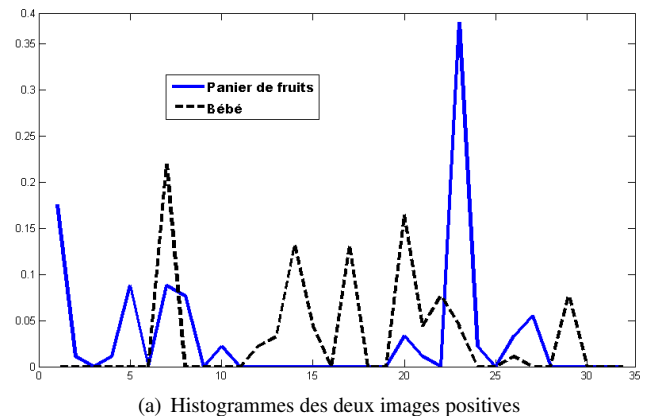


(c) Image négative 1

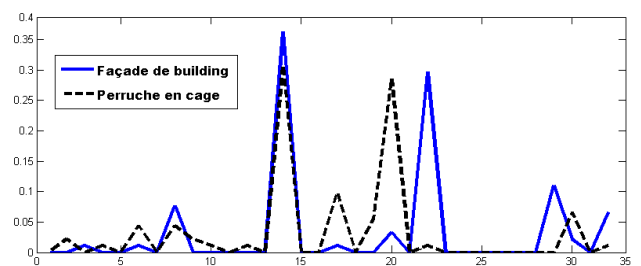


(d) Image négative 2

Figure 4: Quelques images de l'ensemble d'apprentissage



(a) Histogrammes des deux images positives



(b) Histogrammes des deux images négatives

Figure 5: Histogrammes de répartition des coefficients Wave Atoms des images de la figure 4

moyen de réussite sur ces images est de 56,15%. C'est un taux de classification intéressant au vu de la diversité des images et de la complexité de discrimination de l'émotion qu'elles contiennent.

Les meilleurs descripteurs sont les coefficients Wave Atoms, pour notre ensemble de test. L'échelle 4 contient

donc bien suffisamment d'information dans notre cas pour caractériser convenablement les textures.

Les coefficients de l'échelle 4 de la transformée en Wave Atoms sont suivis de très près par l'énergie des caractéristiques de Gabor. Ces deux types de descripteurs décrivent la même information dans l'image : les textures. Cette précision obtenue au moment de la classification est certainement liée au nombre d'orientations utilisé et à la configuration de la gaussienne de convolution qui se rapprochent des modèles proposés pour une description proche de la vision par le cortex visuel.

Après les caractéristiques décrivant la texture, les descripteurs locaux SIFT offrent des résultats satisfaisants. Ce sont des descripteurs assez robustes et intéressants dans l'indexation des images. Ceci justifie sans doute ces résultats intéressants.

On aurait pu penser que la couleur influence énormément sur les émotions et que la classification suivant les couleurs donnerait les meilleurs résultats. Bien évidemment la couleur est un paramètre important dans l'interprétation des émotions mais nos taux de classification bas avec les descripteurs de couleur sont probablement liés à la nature même des descripteurs, sans doute mal choisis, et à leur exploitation. C'est d'ailleurs l'une des familles de descripteurs qu'il serait intéressant d'étudier différemment lors de la poursuite de nos travaux puisque nous sommes convaincus de pouvoir accroître les taux de classification, par une meilleure exploitation des caractéristiques colorimétriques. Les descripteurs globaux GIST sont ceux qui donnent les moins bons résultats lors de la discrimination des images. Il faut garder à l'esprit que le réseau de neurones utilisé s'est employé à régler ses paramètres de façon à avoir un apprentissage parfait. Les images étant diversifiées, on peut comprendre qu'il y ait quelques déconvenues lors de la discrimination.

4.4 Comparaison des résultats obtenus avec des travaux précédents

Il est difficile de comparer les travaux sur l'extraction de l'impact émotionnel entre eux, à cause des différences entre les stratégies de travail et les bases de tests. On peut, néanmoins, évoquer quelques taux de classification obtenus dans les travaux précédents. Par exemple les taux de classification obtenus par Wei et al. [20] sont compris entre 42,3% et 50,25%. Ils proposent une classification émotionnelle des images basée sur des aspects sémantiques de la couleur.

Le taux de réussite de Dellandréa et al. [2] est de 52% sur leur base de tests initiale¹. Pour avoir ces résultats, les caractéristiques utilisées sont les suivantes :

- Les moments de couleur ;
- Les histogrammes de couleur ;
- Les corrélogrammes de couleur ;
- Les caractéristiques texture de Tamura ;

1. Base construite à partir du site <http://www.gettyimages.com/>

- L'histogramme des orientations des lignes et des contours ;
- Les caractéristiques d'harmonie ;
- Les caractéristiques dynamiques.

Ces résultats sont obtenus avec 80% de la base de référence utilisée pour l'apprentissage et 20% pour les tests. Dellandréa et al. [2], ont utilisé les mêmes descripteurs sur la base de IAPS [7]. Pour cette base de données qui est considérée comme une référence pour les études psychologiques, le taux de classification est plus élevé et monte à 64,6%.

La base de IAPS améliore aussi les résultats obtenus par Machajdik et Handurry [11] dont le taux de réussite passe à 68%. Les différentes caractéristiques utilisées sont :

- Caractéristiques colorimétriques ;
- Caractéristiques liées à la texture ;
- Caractéristiques liées à la composition (profondeur des détails) ;
- Caractéristiques liées au contenu (détection de visages, peau).

Au regard de ces différentes stratégies et différentes bases de tests, nos résultats s'avèrent donc intéressants.

5 Conclusions et perspectives

L'extraction de l'impact émotionnel des images nécessite la prise en compte de plusieurs paramètres. Nous avons modélisé au mieux ces paramètres par les attributs suivants :

- Les différentes couleurs présentes dans les images ;
- Les différentes textures des images ;
- Le contenu des images grâce à des descripteurs locaux et globaux.

Nos travaux ont la particularité d'avoir été faits sur une base d'images diversifiée que nous allons mettre à la disposition de la communauté image. Nous avons également utilisé des descripteurs récents et très efficaces. Ils ont permis de déboucher sur des taux de classification très encourageants. Nous obtenions ainsi un taux de réussite moyen de 56,15%. Ces résultats sont intéressants dans la mesure où ils confirment que nos descripteurs sont relativement cohérents par rapport à l'information émotionnelle contenue dans les images.

On pourrait bien sûr tester d'autres classifieurs comme les SVM. Mais l'une des perspectives essentielles à nos travaux serait d'évaluer notre système d'extraction de l'impact émotionnel sur la base de IAPS pour se comparer efficacement aux autres travaux du domaine.

Par ailleurs, les couleurs ont été désignées comme un important vecteur émotionnel dans les images, dans de nombreux travaux que nous avons référencés de manière non exhaustive. Même si aucun résultat unanime n'existe sur le lien exact entre les différentes couleurs et les émotions, elles restent une piste intéressante dans l'extraction de l'impact émotionnel. Il serait logique de continuer nos travaux en explorant une autre modélisation de l'information couleur afin d'en tirer un maximum de caractéristiques discriminantes pour les émotions. Aussi, l'utilisation des Op-

ponentSIFT proposés par van de Sande et Snoek [18] serait plus indiquée puisqu'aucun *a priori* sur la base d'images n'est connu. Ils permettraient également de prendre en compte l'information colorimétrique en même temps que les objets.

L'intégration des différents descripteurs de l'attention visuelle [12] pourrait être également intéressante pour améliorer nos résultats. Ceci pourrait être une piste intéressante à explorer dans le but d'obtenir des résultats conformes à une évaluation humaine. Nous avons à cet effet, prévu d'organiser de nouveaux tests psycho-visuels avec un oculomètre afin d'évaluer l'attention visuelle des observateurs en fonction des différentes émotions.

Références

- [1] Daniel Beresniak. *ABC des couleurs Leurs incidences dans votre vie quotidienne*, pages 11–38. Mars 1990.
- [2] Emmanuel Dellandréa, Ningning Liu, Bruno Tellez, and Liming Chen. Reconnaissance de la sémantique émotionnelle portée par les images. Technical report, Ecole Centrale de Lyon, LIRIS, UMR5205, 25 Nov 2010.
- [3] Laurent Demanet and Lexing Ying. Wave atoms and time upscaling of wave equations. *Numerische Mathematik*, 113 :1–71, 2009.
- [4] Christine Fernandez-Maloigne, Alain Trémeau, and Pierre Bonton. *Image Numérique Couleur : De l'acquisition au traitement*. 2004.
- [5] S.E. Grigorescu, N. Petkov, and P. Kruizinga. Comparison of texture features based on gabor filters. *Image Processing, IEEE Transactions on*, 11(10) :1160 – 1167, oct 2002.
- [6] Naz Kaya and Helen H. Epps. Color-emotion associations : Past experience and personal preference. *AIC Colors and Paints, Interim Meeting of the International Color Association*, 2004.
- [7] P.J. Lang, M.M. Bradley, and B.N. Cuthbert. International affective picture system (IAPS) : Affective ratings of pictures and instruction manual. technical report A-8. Technical report, University of Florida, 2008.
- [8] David G. Lowe. Object recognition from local scale-invariant features. *International Conference on Computer Vision*, 2 :1150–1157, 1999.
- [9] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60 :91–110, 2004.
- [10] Marcel P. Lucassen, Theo Gevers, and Arjan Gijzen. Adding texture to color : quantitative analysis of color emotions. In *CGIV 2010 Final Program and Proceedings*, 2010.
- [11] Jana Machajdik and Allan Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, MM '10, pages 83–92, New York, NY, USA, 2010.
- [12] Olivier Le Meur. *Attention sélective en visualisation d'images fixes et animées affichées sur écran : Modèles et évaluation de performances-Applications*. PhD thesis, École polytechnique de l'Université de Nantes : École doctorale STIM, Octobre 2005.
- [13] Aude Oliva and Antonio Torralba. Modeling the shape of the scene : A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42 :145–175, 2001.
- [14] Li-Chen Ou, M. Ronnier Luo, Andrée Woodcock, and Angela Wright. A study of colour emotion and colour preference. part i : Colour emotions for single colours. *Color Research & Application*, 29(3) :232–240, 2004.
- [15] M. Paleari and B. Huet. Toward emotion indexing of multimedia excerpts. In *Proceedings on Content-Based Multimedia Indexing, International Workshop*, pages 425–432, 2008.
- [16] Simon Rigoulot. *Impact comportemental et électrophysiologique de l'information émotionnelle en vision périphérique*. PhD thesis, Université de Lille II - École Doctorale de Biologie-Santé, Septembre 2008.
- [17] Imen Tayari, Nhan Le Thanh, and Chokri Ben Amar. Modélisation des états émotionnels par un espace vectoriel multidimensionnel. Technical report, Laboratoire Informatique, Signaux et Systèmes de Sophia Antipolis, Décembre 2009.
- [18] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9) :1582–1596, 2010.
- [19] Wei-Ning Wang and Ying-Lin Yu. Image emotional semantic query based on color semantic description. In *The Fourth International Conference on Machine Learning and Cybernetics*, volume 7, pages 4571–4576, 2005.
- [20] Kaiping Wei, Bin He, Tao Zhang, and Wenya He. *Image Emotional Classification Based on Color Semantic Description*, volume 5139 of *Lecture Notes in Computer Science*, pages 485–491. Springer Berlin / Heidelberg, 2008.